

UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA

IN RE GOOGLE GENERATIVE AI  
COPYRIGHT LITIGATION.

Case No. 5:23-cv-03440-EKL (SVK)

**ORDER RE DISCOVERY DISPUTES**

Re: Dkt. Nos. 140, 144, 148

Before the Court are two discovery disputes regarding the selection of Google LLC and Alphabet Inc. (together, “Google”) custodians (Dkt. 144) and the production of Google source code (Dkt. 140). Prior to an omnibus in-person discovery hearing on June 18, 2025, the Court issued rulings and guidance to assist the Parties and the Court in preparing for the hearing (Dkt. 148). The Court commends the Parties for their further meet and confer efforts and the progress made toward resolving these disputes. Remaining issues were addressed by counsel at the hearing.

The Court’s rulings and reasoning for same are as set forth on the record and summarized below for convenience of all concerned.<sup>1</sup>

**I. Custodians**

1. Plaintiffs shall receive an additional 12 Google custodians, for a total of 25 Google custodians, of which these 12 may include either or both of the Google declarant witnesses: Tyler Liechty and Li Xiao (Dkt. 152).
2. Of these 12, the Plaintiffs may identify 9 (including, optionally, the declarants) and Google shall identify the remaining 3.

---

<sup>1</sup> In the event of a discrepancy between the Court’s oral ruling and this written Order, the Order controls.

3. In choosing their Google custodians, each Party must prioritize those that are most likely to impact class certification.
4. Production of documents from Google custodians, as well as from Plaintiff custodians, shall be on a **rolling basis** with results being provided **every two weeks**, to be completed **no later than July 29, 2025**. Given that production is expedited, any privileged material produced qualifies to be “clawed back” under Federal Rule of Evidence 502(b).

## II. Source Code

1. Plaintiffs “ha[ve] the burden of establishing that the disclosure of source code is both relevant and necessary to the action,” (*Synopsys, Inc. v. Nassda Corp.*, 2002 WL 32749138, at \*1 (N.D. Cal. Sept. 16, 2022)), and is not “overbroad [or] disproportionate to the case,” (*Uniloc USA, Inc. v. Apple Inc.*, 2018 WL 2002979, at \*2 (N.D. Cal. Apr. 30, 2018)). At the hearing, Google identified a protocol it has established of creating data cards that allows the Plaintiffs to identify and select datasets of interest for inspection, which Plaintiffs have done. Google represents that these datasets include the content of training data and the content’s associated metadata. Plaintiffs’ remote access to these datasets will be available starting next week, June 23, 2025.
2. Based upon arguments of counsel at the hearing, this protocol should address most if not all of Plaintiffs’ needs for the purposes of class certification. The Court therefore **DENIES without prejudice** each request for source code.
3. If there remain critical data that are unattainable following production and review of selected data cards, then the Parties must meet and confer regarding any remaining, narrow and specific need for source code. If the Plaintiffs are still unable to attain the data needed, they may file a renewed discovery dispute with this Court.

////

////

////

////

**SO ORDERED.**

Dated: June 18, 2025

A handwritten signature in black ink, reading "Susan Van Keulen". The signature is fluid and cursive, with the first name "Susan" and last name "Van Keulen" clearly distinguishable.

---

SUSAN VAN KEULEN  
United States Magistrate Judge